

# **Predicting the “Customer Lifetime Value”**

## **Problem Description**

To predict the Customer lifetime value for an auto insurance company based on different quantitative and qualitative features provided.

Forecasting is an important approach to take an optimal decision and implement appropriate action plans.

A major non-life insurance company wants to evaluate customer lifetime value based on each customer's demographics and policy information including claim details. The CLV is a profitability metric in terms of a value placed by the company on each customer and can be conceived in two dimensions: the customer's present Value and potential future Value.

You are expected to create an analytical and modelling framework to predict the lifetime value of each customer based on the quantitative and qualitative features provided in the dataset and also cluster the train dataset to understand behaviour of each segment.

### **Files**

- i. Train.csv (Build model)
- ii. Test.csv ( Model will be tested on this in grader tool).
- iii. Sample\_Submission: Customer ID in the order in which the predictions need to be submitted and randomly generated sample predictions.

### **Main Tasks:**

- Data Pre-processing/Exploratory Data Analysis/Visualization on Train Data.
- Model Building
- Clustering and insights from clustering your data. This is left open-ended but you are expected to come up with your own ideas on which data subset is worth clustering and what insights you can get out of it.
- Presentation and Viva

### **Submission 1 (24<sup>th</sup> Feb 2018 by 6 pm):**

1. Project report
2. Code file
3. Prediction file

### **Submission 2 (25<sup>th</sup> Feb 2018 by 6 am):**

**Final presentation for viva with the name “viva.ppt”**

**Primary Accuracy Metric:**  $R^2$  (Grader evaluation will calculate this metric with baseline on TRAIN data)